

Visualizing Evolvability with Price's Equation

Mitchell A. Potter

Naval Research Laboratory
Washington, DC 20375
mpotter@aic.nrl.navy.mil

Jeffrey K. Bassett

Naval Research Laboratory
Washington, DC 20375
bassett@aic.nrl.navy.mil

Kenneth A. De Jong

George Mason University
Fairfax, VA 22030
kdejong@cs.gmu.edu

Abstract- The term “premature convergence” has been used for many years as an explanation as to why an evolutionary algorithm fails to find a global optimum, without providing much insight into how to fix the problem and/or avoid it in the future. In this paper we tie these issues to notions of (lack of) evolvability that have been explored in the population genetics community for many years. In particular, we show how the central equation in Price's Theorem can be extended in such a way as to separate out the individual contributions that reproductive operators make to evolvability, paving the way for better designed EAs in the future.

1 Introduction

The term “premature convergence” has been used for many years in the evolutionary computation community to describe situations in which an evolutionary optimization algorithm is caught in a local basin of attraction and, as such, has little chance of further progress towards a global optimum. A common remedy is to try to find a better balance between “exploration and exploitation” which, for complex and poorly understood objective landscapes, is a difficult and challenging task. A standard approach is to do some EA parameter tuning over multiple runs in an attempt to find such a balance. A more satisfying approach would be to have an EA capable of dynamically adapting this balance.

In order to achieve this we need some internal mechanism for measuring this balance. A classic example of this is the “1:5 rule” used in most ES EAs to dynamically adapt the step size of the Gaussian mutation operator (Schwefel 1981). The focus of this paper is on whether there are more general mechanisms for maintaining an effective exploration/exploitation balance. In this respect we are indebted to Lee Altenberg who, as early as 1994 was telling the EC community to pay more attention to Price's Theorem (Altenberg 1994).

In the population genetics community these issues are discussed in terms of the “evolvability” of a population, that is, its ability to continue to improve over time with respect to some measure of quality. Hence, our notion of premature convergence closely resembles their notion of loss of evolvability. Further, results such as Price's Theorem suggest ways of decomposing EA dynamics so as to see the individual contributions of selection and reproduction to evolvability.

In this paper we describe some preliminary work in which we extend the central equation of Price's Theorem

in a new and interesting way to help visualize EA dynamics as they relate to evolvability. The ultimate goal is to use these techniques to design better EAs.

2 Background

In 1970, George Price published the article *Covariance and Selection* (Price 1970) in which he presented an equation that has proved to be a major contribution to the field of evolutionary genetics. His ultimate goal was to create a general theory of selection. He began by addressing the issue of selection in biology and evolution, and used his equation to describe and predict the change in gene frequency from one generation to another.

In particular he noticed that a covariance relationship exists between the number of successful offspring that an individual produces (this is the biological notion of fitness) and the frequency of any given gene in that individual. If this covariance value is high, then the existence of that gene is a good predictor of selection.

Steven Frank (1995) summed up Price's contributions with this comment: “The brilliance of the Price Equation is that it adds nothing to the fundamental simplicity of evolutionary change but, by making a few minor rearrangements and changes in notation, the equation provides an easier and more natural way to reason about complex problems.”

2.1 Price's Equation

In this section we will provide a brief explanation of the central equation of Price's Theorem. Given the following:

- P_1 is the parent population
- P_2 is the child population
- Q_1 is some measured attribute (e.g. gene frequency) averaged over population P_1
- Q_2 is the same measured attribute averaged over population P_2

Price's Equation states that

$$\Delta Q = \frac{\text{Cov}(z, q)}{\bar{z}} + \frac{\sum z_i \Delta q_i}{N\bar{z}}, \quad (1)$$

where,

- $\Delta Q = Q_2 - Q_1$
- N is the number of individuals in P_1

- z_i is the number of children to which parent i contributed genetic material
- $\bar{z} = \frac{\sum_i z_i}{N}$
- q_i is the measurement of some attribute of parent i such as the number of occurrences of a particular gene or combination thereof
- q'_i is the average value of the q_i attribute measured in the children of parent i
- $\Delta q_i = q'_i - q_i$

In summary, Price's Equation estimates the change in a measurable attribute from the parent population to the child population. Furthermore, the equation separates the change attributable to selection from the change attributable to the genetic operators.

2.2 Applications in Evolutionary Computation

The most direct application of Price's work involves situations in which the attribute being measured is gene frequency. For example, Langdon and Poli (2002) show how measuring gene frequencies is equivalent to determining the frequency of use of the available primitives in the evolving solution trees. They were able to use this information to diagnose the probable causes of poorer performing runs.

However, Price's ultimate goal was to create a general theory of selection, not just as it applies to biology. Although its initial application was to modeling changes in gene frequency, he clearly states "It can also be applied to non-genetical selection" and gives an example in his paper of applying it to student IQs to determine which students will be likely to pass a class.

Altenberg (1995) also demonstrates that gene frequency is not the only attribute of the individuals which can be measured. He identifies several different measurement functions which could be useful, including mean fitness from both the biological and evolutionary computation perspectives, frequency of schemata, and evolvability.

We are particularly interested in using Price's Equation to help us better understand the dynamics of evolvability—the ability of an EA to continue to make improvements in fitness over time. The changes in fitness from one generation to the next is an important metric in the field of quantitative genetics (Mühlenbein 1998). If we focus on changes in fitness as a measurement function as follows:

$$\Delta Q = \bar{f}(t+1) - \bar{f}(t), \quad (2)$$

where $\bar{f}(t)$ is the average fitness of a population at generation t , then Price's Equation tells us how ΔQ can be decomposed in a useful manner.

Since Price's Equation is highly theoretical, we need to validate that equation 1 is a reasonably good estimator of the observed change in fitness between parents and children of real EAs (equation 2). Figure 1 gives an example of this by plotting the results, averaged over 100 runs, of applying Price's Equation to a running evolutionary system

optimizing the Schwefel function that will be described in Section 4.1. The solid line indicates the average observed change in fitness between the parents and children from one generation to the next, and the dots indicate the value of ΔQ calculated using Price's Equation. From the plot we can see that in this particular example ΔQ tracks the actual observed change in fitness almost perfectly. This was also validated for the Hierarchical If-and-Only-If function described in Section 4.2.

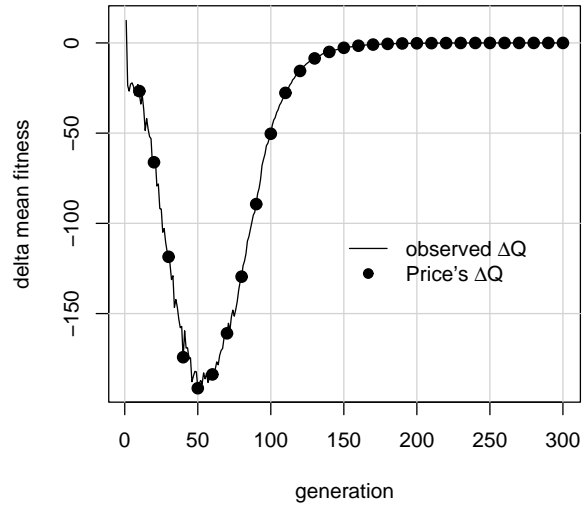


Figure 1: Comparison of Price's ΔQ and the measured change in fitness between parents and children

3 Extending Price's Equation

If we are to achieve our goal of understanding how evolvability is affected by individual reproductive operators, we need to extend Price's Equation to further separate the effects of the individual reproductive operators as follows:

$$\Delta Q = \frac{\text{Cov}(z, q)}{\bar{z}} + \sum_{j=1}^k \frac{\sum_i z_i \Delta q_{ij}}{N\bar{z}}, \quad (3)$$

where k is the number of genetic operators, q'_{ij} is the average value of the measurement function applied to the children of i after the application of operator j , and $\Delta q_{ij} = q'_{ij} - q'_{i(j-1)}$. Also note that $q'_{i0} = q_i$ and $q'_{ik} = q'_i$.

A proof of equation 3 consists of showing that the summation term is equivalent to the summation term of Price's original equation as follows:

$$\begin{aligned} \sum_{j=1}^k \frac{\sum_i z_i \Delta q_{ij}}{N\bar{z}} &= \frac{1}{N\bar{z}} \sum_{j=1}^k \sum_i z_i \Delta q_{ij} \\ &= \frac{1}{N\bar{z}} \sum_{j=1}^k \sum_i z_i (q'_{ij} - q'_{i(j-1)}) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N\bar{z}} \sum_{j=1}^k \left(\sum_i z_i q'_{ij} - \sum_i z_i q'_{i(j-1)} \right) \\
&= \frac{1}{N\bar{z}} \left(\sum_i z_i q'_{ik} - \sum_i z_i q'_{i0} \right) \\
&= \frac{1}{N\bar{z}} \left(\sum_i z_i (q'_{ik} - q'_{i0}) \right) \\
&= \frac{\sum_i z_i \Delta q_i}{N\bar{z}}
\end{aligned}$$

We illustrate the utility of the extended version of Price's Equation in the following sections.

4 Visualization Results

We have applied the extended Price's Equation described in Section 3 to two function optimization problems, with very different characteristics from one another, in order to see whether this technique helps us better visualize and understand the dynamics of a running evolutionary algorithm. The measurement function we use is mean fitness computed directly from the objective function.

In all of the experiments described below the evolutionary algorithms are implemented with the ECKit Java class library developed by Potter (1998). The population size is 1000, ranked selection is used, and the representation is either binary or real-valued as noted in the problem descriptions below. When a binary representation is used, a bit-flipping mutation is applied at the rate of $1/L$, where L is the length of the chromosome. When a real-valued representation is used, a Gaussian mutation with adaptive standard deviations is applied to each gene as described in (Bäck and Schwefel 1993). With both representations, two-point crossover is applied at a rate of 0.6. Unless otherwise noted, experiments are performed for 100 runs and the results averaged to produce the graphs shown in this section.

4.1 Real-Valued Function Optimization

We begin with a standard problem from the function optimization literature introduced by Schwefel (1981). The objective function

$$f(\vec{x}) = 418.9829n + \sum_{i=1}^n x_i \sin\left(\sqrt{|x_i|}\right)$$

defines a landscape covered with a lattice of large peaks and basins. The predominant characteristic of the function is the presence of a second-best minimum far away from the global minimum, intended to trap optimization algorithms on a suboptimal peak. The best minimums are near the corners of the space. As in (Potter 1997) we have added the term $418.9829n$ to the Schwefel function so its global minimum will be zero, regardless of dimensionality. In our experiments the problem has thirty independent variables constrained to the range $(-500.0, 500.0)$.

We are interested in comparing the dynamics of evolving solutions to the Schwefel function with real-valued versus binary representations. As shown in Figure 2, standard

best fitness curves indicate that the binary representation approaches the global minimum more quickly, but give us little insight into how the operators are exploring or exploiting the fitness landscape.

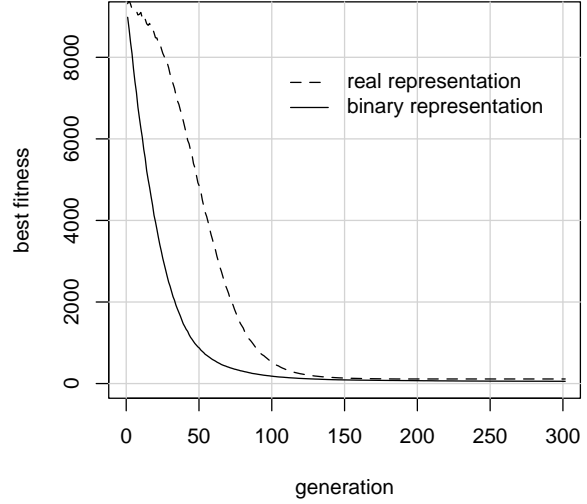


Figure 2: Comparison of best fitness curves from Schwefel function optimization using real-valued versus binary representations

Figures 3 and 4 shows the results from applying the extended Price's Equation, where the balance between the exploitation of ranked selection and the exploration of mutation can be clearly visualized. Note that the Schwefel function is being minimized, so exploitation is manifested as a negative change in mean fitness from one generation to the next, while exploration is manifested as a positive (disruptive) delta mean fitness. With respect to the binary representation, Figure 3 shows that the mutation operator produces significant exploration to the end, thus maintaining the possibility of discovering a better solution, while the real-valued representation with adaptive mutation shown in Figure 4 anneals the standard deviations to the point where evolution is essentially shut down by generation 150. This graph clearly shows that the real-valued evolutionary algorithm might as well be restarted at this point if a sufficient solution has not already been found.

Turning our attention to the crossover curves, in both figures it appears that this operator has little effect in solving this particular problem. However, if we focus on a single run rather than the average of 100 runs, and expand the y-axis to better see small differences as in Figure 5, crossover does indeed appear to be contributing to the evolution of a solution as it distributes both good and bad chromosome segments through the population. This is exhibited by the positive and negative fluctuations in the crossover curve.

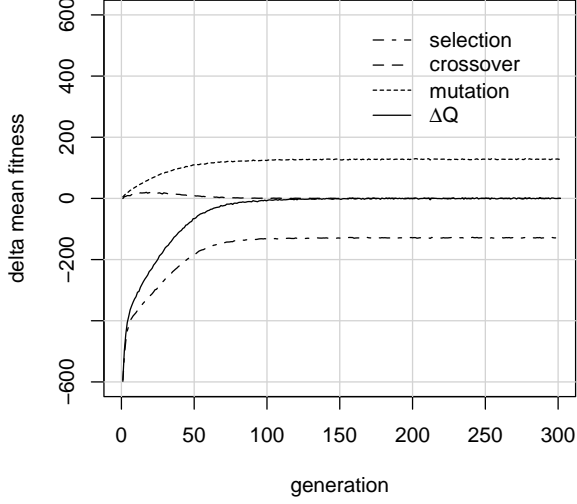


Figure 3: Effect of selection and genetic operators on Schwefel function optimization using binary representation

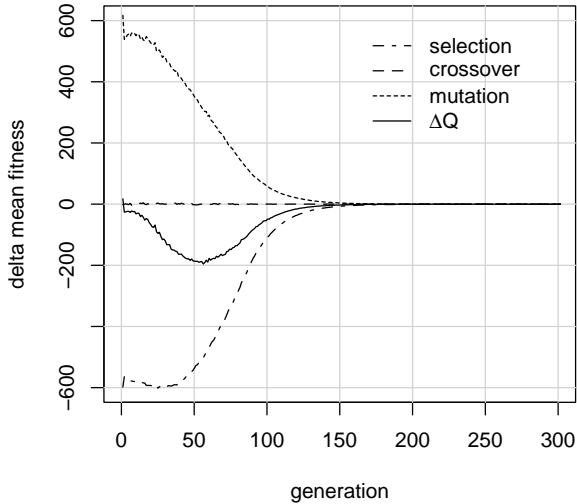


Figure 4: Effect of selection and genetic operators on Schwefel function optimization using real-valued representation

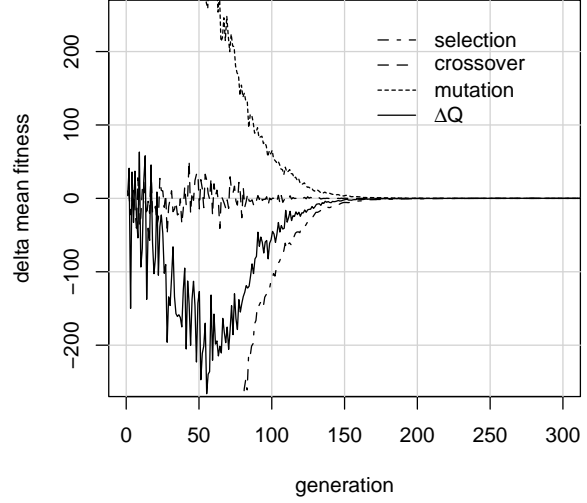


Figure 5: Single-run effect of selection and genetic operators on Schwefel function optimization using real-valued representation

4.2 Building-Block Problems

To more fully visualize the effect of crossover, we applied our extended Price’s Equation to the optimization of the Hierarchical If-and-Only-If function (HIFF) developed by Watson (2001). The HIFF function is in the class of building-block problems that are designed specifically to be exploitable by the crossover operator while intentionally being poorly matched to a symbol-flipping mutation operator. The HIFF objective function is as follows:

$$f(S) = \begin{cases} 1 & \text{if } |S| = 1 \\ |S| + f(S_L) + f(S_R) & \text{if } |S| > 1 \wedge \forall_{ij}: s_i = s_j \\ f(S_L) + f(S_R) & \text{otherwise,} \end{cases}$$

where S is a string of symbols, S_L is the left half of the string, and S_R is the right half of the string. Given a string of length 2^k , HIFF is characterized by a hierarchy of $k + 1$ levels. Each level consists of 2^{k-j} substrings of length j , where j ranges from 0 to k , and each substring only contributes to the value of the objective function when all its symbols are identical. These internally consistent substrings form “building blocks” that can be recombined by crossover to form value conferring strings at a higher level in the hierarchy. Binary strings of length 2^6 are used in our experiments, which produce a maximum possible objective function value is 448 and a minimum possible value is 64.

We contrast the extended Price’s Equation to an ablation study, a visualization technique in use for decades in which some genetic operators are turned off during an experiment to more clearly see the effect of the remaining ones on the dynamics of the evolutionary system (De Jong 1975; Spears 2000). As shown in Figure 6, performing the ablation study by selectively turning off crossover and mutation demonstrates that crossover is indeed a useful operator for opti-

mizing the HIFF function. Not only does crossover alone find a better solution on average than mutation alone, but crossover by itself also finds a better solution on average than both crossover and mutation used together!

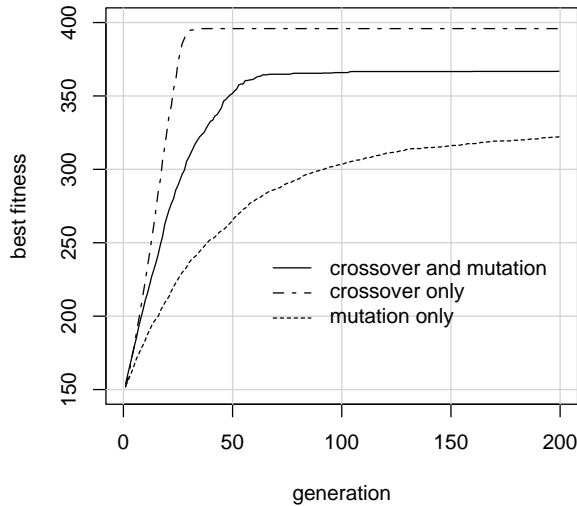


Figure 6: Comparison of best fitness curves from HIFF function optimization using crossover and mutation versus crossover only and mutation only

To gain a better understanding of the evolutionary dynamics in play here, we apply the extended Price’s Equation to HIFF optimization using both crossover and mutation as shown in Figure 7, where it is possible to visualize some of the dynamics of the interaction between crossover and mutation that is lost in the ablation study. Specifically, mutation feeds crossover disruptive genes which are spread through the population when crossover recombines them with other individuals. This is exhibited in the figure by the slightly negative (disruptive) recombination curve throughout the run.

We can also combine Price’s Equation with ablation by turning off mutation and measuring the effect of only crossover and selection as shown in Figure 8. This figure clearly shows the strong initial flurry of activity as crossover combines internally consistent substrings into larger and larger fitness conferring strings, and the rapidly converging population as manifested by the sharply dropping exploitive effect of selection prior to generation 50.

5 Conclusions and Future Work

Our goal is to be able to design EAs that are capable of maintaining an effective exploration/exploitation balance. Our approach is to leverage off existing work in the population genetics community relating to the ability of a population to maintain its ability to evolve over time. In particular, we have shown how the central equation in Price’s Theorem can be extended to separate out the contributions

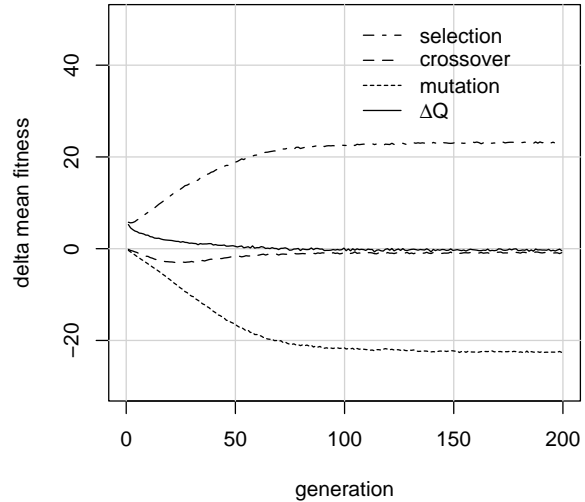


Figure 7: Effect of selection, crossover, and mutation on HIFF function optimization

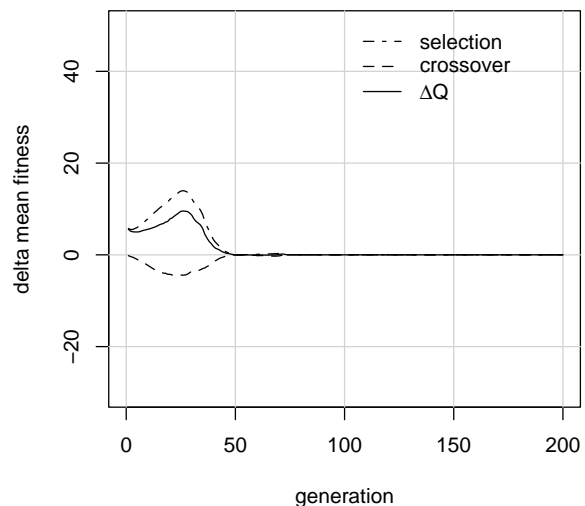


Figure 8: Effect of selection and crossover on HIFF function optimization when mutation is deactivated

to evolvability made by individual reproductive operators.

Using this extension we have developed a visualization tool that allows us to visualize the balance between exploration and exploitation in new and interesting ways. In particular, we can see the different roles that selection, mutation, and crossover play, and how they interact with one another to maintain a reasonable balance between the disruptive effect of exploration and the convergent effect of exploitation.

We have explored a single measurement function—mean fitness computed directly from the objective function. Other measurement functions may give us additional insights. Our next step is to explore some different measurement functions and use the insights from these experiments to help us design operators and representations better matched to the problems we are trying to solve. Ultimately, we would like to use these results to provide an EA with the feedback it needs to dynamically adjust the exploration/exploitation balance.

Acknowledgments

We thank Donald Sofge, Magdalena Bugajska, Myriam Abramson, and Paul Wiegand for helpful discussions on the topic of Price's Equation. The work reported in this paper was supported by the Office of Naval Research under work request N0001403WX20212.

Bibliography

- Altenberg, L. (1994). The evolution of evolvability in genetic programming. In K. E. K. Jr. (Ed.), *Advances in Genetic Programming*, Chapter 3, pp. 47–74. M.I.T. Press.
- Altenberg, L. (1995). The schema theorem and Price's theorem. In L. D. Whitley and M. D. Vose (Eds.), *Foundations of Genetic Algorithms III*, pp. 23–49. Morgan Kaufmann.
- Bäck, T. and H.-P. Schwefel (1993). An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation* 1(1), 1–23.
- De Jong, K. A. (1975). *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. Ph. D. thesis, University of Michigan.
- Frank, S. A. (1995). George price's contributions to evolutionary genetics. *Journal of Theoretical Biology* 175, 373–388.
- Langdon, W. B. and R. Poli (2002). *Foundations of Genetic Programming*. Berlin Heidelberg: Springer-Verlag.
- Mühlenbein, H. (1998). The equation for response to selection and its use for prediction. *Evolutionary Computation* 5(3), 303–346.
- Potter, M. A. (1997). *The Design and Analysis of a Computational Model of Cooperative Coevolution*. Ph. D. thesis, George Mason University.
- Potter, M. A. (1998). Overview of the evolutionary computation toolkit. <http://cs.gmu.edu/~mpotter/>.
- Price, G. (1970). Selection and covariance. *Nature* 227, 520–521.
- Schwefel, H.-P. (1981). *Numerical optimization of Computer models*. Chichester: John Wiley & Sons, Ltd.
- Spears, W. M. (2000). *Evolutionary Algorithms: the role of mutation and recombination*. Berlin: Springer.
- Watson, R. A. (2001). Analysis of recombinative algorithms on a non-separable building-block problem. In W. N. Martin and W. M. Spears (Eds.), *Foundations of Genetic Algorithms VI*, pp. 69–89. Morgan Kaufmann.